Accepted Manuscript

Bayesian inference with subset simulation: Strategies and improvements

Wolfgang Betz, Iason Papaioannou, James L. Beck, Daniel Straub

PII: S0045-7825(17)30721-1
DOI: https://doi.org/10.1016/j.cma.2017.11.021
Reference: CMA 11680

To appear in: *Comput. Methods Appl. Mech. Engrg.*

Received date : 10 May 2017
Revised date : 8 November 2017
Accepted date : 13 November 2017

Please cite this article as: W. Betz, I. Papaioannou, J.L. Beck, D. Straub, Bayesian inference with subset simulation: Strategies and improvements, *Comput. Methods Appl. Mech. Engrg.* (2017), https://doi.org/10.1016/j.cma.2017.11.021

Bayesian inference with Subset Simulation: strategies and improvements

Wolfgang Betz[a,*], Iason Papaioannou[a], James L. Beck[b], Daniel Straub[a]

[a] *Engineering Risk Analysis Group, Technische Universität München, 80333 München, Germany*
[b] *Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125*

Abstract

Bayesian Updating with Structural reliability methods (BUS) reinterprets the Bayesian updating problem as a structural reliability problem; i.e. a rare event estimation. The BUS approach can be considered an extension of rejection sampling, where a standard uniform random variable is added to the space of random variables. Each generated sample from this extended random variable space is accepted if the realization of the uniform random variable is smaller than the likelihood function scaled by a constant c. The constant c has to be selected such that $1/c$ is not smaller than the maximum of the likelihood function, which, however, is typically unknown a-priori. A c chosen too small will have negative impact on the efficiency of the BUS approach when combined with sampling-based reliability methods. For the combination of BUS with Subset Simulation, we propose an approach, termed aBUS, for *adaptive* BUS, that does not require c as input. The proposed algorithm requires only minimal modifications of standard BUS with Subset Simulation. We discuss why *aBUS* produces samples that follow the posterior distribution – even if $1/c$ is selected smaller than the maximum of the likelihood function. The performance of *aBUS* in terms of the computed evidence required for Bayesian model class selection and in terms of the produced posterior samples is assessed numerically for different example problems. The combination of BUS with Subset Simulation (and aBUS in particular) is well suited for problems with many uncertain parameters and for Bayesian updating of models where it is computationally

*Corresponding author
Email addresses: wolfgang.betz@tum.de (Wolfgang Betz), iason.papaioannou@tum.de (Iason Papaioannou), jimbeck@caltech.edu (James L. Beck), straub@tum.de (Daniel Straub)

demanding to evaluate the likelihood function.

Keywords: Bayesian updating, Bayesian model class selection, Subset Simulation, structural reliability, MCMC, BUS

1. Introduction

Bayesian inference provides a consistent framework to reduce uncertainties in existing models with new information. Uncertainty is represented by a probability distribution over the model parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^M$. The information about $\boldsymbol{\theta}$ already acquired in the past is described by the *prior* distribution $p(\boldsymbol{\theta})$, which represents one's *initial belief* on the parameters $\boldsymbol{\theta}$. Information/data $\mathbf{d}$ that becomes available in form of *measurements or observations* is embedded in the Bayesian analysis through the likelihood function $L(\boldsymbol{\theta}|\mathbf{d}) = p(\mathbf{d}|\boldsymbol{\theta})$, which comes from substituting $\mathbf{d}$ into a stochastic model that predicts what the data should be for given $\boldsymbol{\theta}$. The learning process in Bayesian inference is formalized through *Bayes' theorem* as:

$$p(\boldsymbol{\theta}|\mathbf{d}) = {c_\mathrm{E}}^{-1} \cdot L(\boldsymbol{\theta}|\mathbf{d}) \cdot p(\boldsymbol{\theta}) \tag{1}$$

where $p(\boldsymbol{\theta}|\mathbf{d})$ is the posterior distribution that represents the posterior state of knowledge about the uncertain parameter vector $\boldsymbol{\theta}$, and c_E is a normalizing scalar.

Except for some special cases, the posterior distribution cannot be derived analytically, and posterior samples are usually generated numerically. Markov chain Monte Carlo (MCMC) methods constitute a popular class of methods to sample from the posterior distribution [1, 2]. One problem of MCMC methods is that after an initial burn-in phase, the samples may not yet have reached the stationary distribution of the Markov chain [3]. That is, finding an appropriate burn-in period in MCMC is often a non-trivial problem. Another issue is that standard MCMC algorithms usually cannot be applied efficiently for problems with many uncertain parameters. Some specialized MCMC algorithms [4, 5, 6, 7] can cope with high dimensional problems, however, they require additional evaluations of the likelihood function or its gradient for each generated sample.

The constant c_E in Eq. (1) is a measure for the plausibility of the assumed model class [8]:

$$c_E = \int_{\Theta} L(\theta|\mathbf{d}) \cdot p(\theta)\, d\theta \tag{2}$$

c_E is referred to as the *evidence* [9]; alternatively it is also known as *marginal likelihood* or *integrated likelihood*. The evidence c_E is required for Bayesian model class selection and model averaging [9, 10, 11]. It is typically challenging to compute the evidence c_E because of the multi-dimensional integral in Eq. (2). If the system is globally identifiable [12, 13], asymptotic approximations [9, 14] can be applied. Otherwise, the evidence is usually estimated numerically. A review of techniques to compute c_E is given in Cheung and Beck [15]. Methods to generate posterior samples are not necessarily suitable to estimate the evidence, and vice versa.

A recently introduced framework for Bayesian updating, called BUS [16], converts the evaluation of posterior densities into an equivalent reliability problem. In structural reliability, probabilities of rare events are estimated [17, 18, 19]. By interpreting the Bayesian updating problem as a rare event estimation, existing structural reliability methods can be used to perform the Bayesian analysis. Moreover, an estimate for the evidence c_E is obtained as a by-product of BUS. The Subset Simulation (SuS) algorithm [20] is a structural reliability method that is well suited for BUS: (i) SuS can efficiently handle problems with many uncertain parameters; (ii) SuS can efficiently estimate very small probabilities that may arise within BUS. The use of SuS in BUS is referred to as *BUS-SuS* in the following.

A limitation of the original BUS is that prior to the analysis a constant c has to be selected [21]: On the one hand, c^{-1} should not be smaller than the maximum value that the likelihood function can take: $c^{-1} \geq L_{\max}$. On the other hand, selecting c^{-1} conservatively large decreases the efficiency of the method. Therefore, an appropriate choice of c is crucial. However, in many cases the maximum of the likelihood function is not known in advance. In some cases, probabilistic information on the optimal value of c can be derived as a function of the data size [16]. For the use of SuS within BUS, two strategies that avoid selecting c have been proposed recently. In [22] an adaptive strategy to learn the maximum of the likelihood function during the simulation

is suggested. In [21] the equivalent structural reliability problem is redefined such that the stopping criterion of SuS depends on c^{-1}, but not the underlying limit-state function. In [23, 24] an alternative strategy to BUS is presented, based on the concept of Approximate Bayesian Computation that also allows the use of SuS for Bayesian updating; it avoids the issue of selecting any constant like c but at the expense of getting only approximate posterior samples. The BUS approach is combined in [25] with an adaptive neural network surrogate model.

This contribution focuses on the application of SuS within BUS. We pick up and extend the idea originally proposed in [22], of learning the constant c^{-1} on the fly. The proposed algorithm is termed *aBUS* (as a substitute for adaptive BUS) and requires only minimal modifications of standard BUS with SuS. We also discuss why *aBUS* produces samples that follow the posterior distribution – even if c^{-1} is selected smaller than the maximum of the likelihood function.

The structure of the paper is as follows: In Section 2, we formally introduce the BUS approach. In Section 3, the combination of BUS and SuS is explained in-depth. The proposed algorithm *aBUS* that adaptively learns the value of c^{-1} is introduced in Section 4. In Section 5, illustrative applications are presented to demonstrate the efficiency of the proposed method numerically using different examples. Section 6 briefly summarizes the obtained findings.

2. Bayesian updating with structural reliability methods

2.1. The idea behind BUS

Straub and Papaioannou show in [16] that a Bayesian updating problem can be interpreted as a structural reliability problem. The principal idea behind BUS (Bayesian Updating with Structural reliability methods) is to add an additional uniformly distributed random variable $\pi \in \Pi = [0, 1]$ to the space of random variables spanned by $\boldsymbol{\Theta}$. The updating problem is then expressed as a structural reliability problem in the augmented random variable space $\boldsymbol{\Theta} \times \Pi$. The "failure" domain Ω of this reliability problem is defined as:

$$\Omega = \{\pi \leq c \cdot L(\boldsymbol{\theta}|\mathbf{d})\} \tag{3}$$

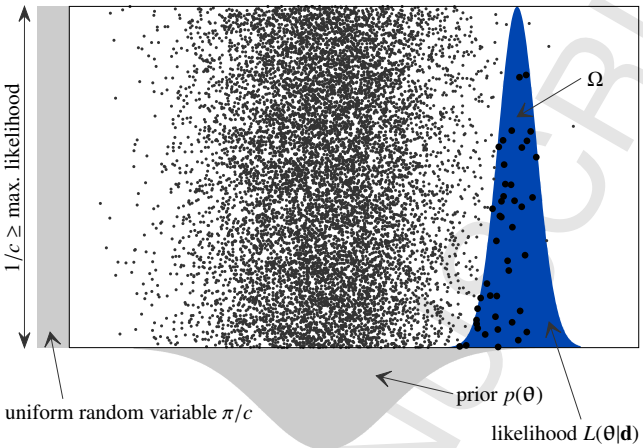


Figure 1: Illustration of the principle of BUS combined with rejection sampling. The highlighted region is the domain Ω defined in Eq. (3). The limit-state function $g(\theta, \pi)$ introduced in Eq. (4) is smaller or equal than *zero* within Ω (it is *zero* at the boundary of Ω), and larger than *zero* outside of Ω. Samples "below" the likelihood (i.e., the samples contained in Ω) are independent samples from the posterior distribution. In this example, 43 out of 10^4 samples are accepted.

where c is a positive constant chosen such that $c \cdot L(\theta|\mathbf{d}) \leq 1$ is maintained for all θ. The domain Ω is illustrated in Fig. 1. Note that Ω denotes both the failure domain and the corresponding event. The link between the domain Ω and the actual Bayesian updating problem is: Samples from the prior distribution of θ that are in Ω follow the posterior distribution [16]. In reliability analysis, the limit-state function is defined such that: $g(\theta, \pi) \leq 0$ if $[\theta, \pi] \in \Omega$; and $g(\theta, \pi) > 0$ if $[\theta, \pi]$ is outside of Ω (see Fig. 1). The limit-state function $g(\theta, \pi)$ that describes the "failure" domain Ω defined in Eq. (3) can be expressed as:

$$g(\theta, \pi) = \pi - c \cdot L(\theta|\mathbf{d}) \tag{4}$$

Optimally, the constant c^{-1} should be chosen as the maximum of the likelihood function, denoted $L_{\max}$ [16]. However, $L_{\max}$ is not always known in advance. In such cases, it is difficult to select c appropriately. An efficient strategy based on *BUS-SuS* that renders a prior selection of c unnecessary is developed in Section 3.

The major appeal of BUS is that any structural reliability method can be used to tackle the associated Bayesian inference problem. The most straight-forward (and sim-

plest) application of BUS is rejection sampling – which corresponds to crude Monte Carlo simulation in the context of structural reliability. The principle of BUS with rejection sampling is illustrated in Fig. 1. The combination of BUS with SuS (see for example [16, 21, 22, 26, 27]) is of particular interest, because it is efficient for very small failure probabilities and its performance does not depend on the dimension M of the vector of uncertain model parameters $\boldsymbol{\theta}$. In Section 3, the combination of BUS and SuS (*BUS-SuS*) is explained and modifications to learn the constant c on the fly are proposed.

2.2. Estimating the evidence in BUS

An estimate for the evidence c_{E} is obtained as a by-product of BUS. Let p_Ω be the probability that samples $[\boldsymbol{\theta}, \pi]$ from the prior distribution fall into Ω, i.e.:

$$p_\Omega = \Pr[\Omega] = \Pr[g(\boldsymbol{\theta}, \pi) \leq 0] \tag{5}$$

p_Ω is the target quantity of interest in a reliability analysis and referred to as the *probability of failure*. In BUS, p_Ω is directly linked to the evidence c_{E} through c^{-1} [16]:

$$c_{\mathrm{E}} = c^{-1} \cdot p_\Omega \tag{6}$$

Note that some reliability methods allow evaluating uncertainty bounds for the estimate of p_Ω. In this case, the statistical uncertainty in the estimated evidence c_{E} can be quantified directly, as the evidence is directly proportional to p_Ω.

2.3. Outline of a simple proof of BUS

A simple proof that demonstrates the validity of BUS is as follows [16]. The quantity $c \cdot L(\boldsymbol{\theta}|\mathbf{d})$ can be expressed as:

$$c \cdot L(\boldsymbol{\theta}|\mathbf{d}) = \int_{0 \leq \pi \leq c \cdot L(\boldsymbol{\theta}|\mathbf{d})} \mathrm{d}\pi \tag{7}$$

Consequently, $L(\boldsymbol{\theta}|\mathbf{d}) \cdot p(\boldsymbol{\theta})$ can be expressed as:

$$L(\boldsymbol{\theta}|\mathbf{d}) \cdot p(\boldsymbol{\theta}) = c^{-1} \int_{0 \leq \pi \leq c \cdot L(\boldsymbol{\theta}|\mathbf{d})} p(\boldsymbol{\theta})\, \mathrm{d}\pi \tag{8}$$

By inserting Eq. (8) into Eq. (2) one can easily prove the validity of Eq. (6), and by inserting Eq. (8) into Eq. (1), one can observe that sampling from the posterior has been converted to sampling from the failure domain of a reliability problem.

2.4. BUS in standard Normal space

For some reliability methods it is convenient to transform the reliability problem to the so-called underlying *standard Normal space*. Let $\mathbf{u}^*$ be a M-dimensional vector whose M components are independent standard Normal random variables. Without loss of generality, we assume that there exists a transformation $\mathrm{T}^*(\cdot)$ such that $\theta = \mathrm{T}^*(\mathbf{u}^*)$. For example, if the joint prior distribution of $p(\theta)$ is known, the mapping can be defined in terms of the Rosenblatt transformation [28]. If the probabilistic description of θ is only available in terms of marginal distributions and correlations, the joint distribution is usually modeled by a Gaussian copula (also known as the Nataf distribution) and the mapping can be achieved through a marginal transformation [29].

Let $\mathbf{u}$ be a $(M+1)$-dimensional vector that extends $\mathbf{u}^*$ by one dimension. For the last component of $\mathbf{u}$, denoted u_{M+1}, one can write: $\pi = \Phi(u_{M+1})$, where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard Normal distribution; i.e. u_{M+1} also follows a standard Normal distribution. Thus, the limit-state function $g(\theta, \pi)$ defined in Eq. (4) can be expressed equivalently as

$$g(\theta, \pi) = G(\mathbf{u}) = \Phi(u_{M+1}) - c \cdot L(\mathrm{T}^*(\mathbf{u}^*)|\mathbf{d}) \tag{9}$$

with $\mathbf{u} = [\mathrm{T}^{*-1}(\theta), \Phi^{-1}(\pi)]$. Note that the prior distribution of $\mathbf{u}$ is $p(\mathbf{u}) = \prod_{i=1}^{M+1} \varphi(u_i)$, where $\varphi(\cdot)$ is the probability density function of the standard Normal distribution.

2.5. BUS with rejection sampling

The most trivial application of BUS results in the *rejection sampling* algorithm [16, 30] (see *Algorithm (1)* below). This algorithm repeatedly proposes a sample $[\tilde{\theta}, \tilde{\pi}]$ from the prior distribution and accepts the sample if it is located in the "failure" domain; i.e., if $[\tilde{\theta}, \tilde{\pi}] \in \Omega$. The accepted sample $\tilde{\theta}$ is a sample from the posterior distribution. The algorithm is repeated until K posterior samples are generated. The posterior samples resulting from the rejection sampling algorithm are statistically independent.

Algorithm (1): rejection sampling

As input the algorithm requires:

– K, the total number of samples to draw from the posterior distribution.

– c, selected such that $c^{-1} \geq L_{\max}$.

The algorithm evaluates the evidence c_{E} and returns K uniformly weighted and statistically independent posterior samples $\boldsymbol{\theta}_{(k)}$ with $k = 1, \ldots, K$.

1. Initialize counters $k = 1$ and $n = 0$.
2. **while** $(k \leq K)$ **do**:
   (a) Propose sample $[\tilde{\boldsymbol{\theta}}, \tilde{\pi}]$:
      i. Draw $\tilde{\boldsymbol{\theta}}$ from the prior distribution $p(\boldsymbol{\theta})$.
      ii. Draw $\tilde{\pi}$ from the uniform distribution that has support $[0, 1]$.
   (b) **if** $\left(g(\tilde{\boldsymbol{\theta}}, \tilde{\pi}) \leq 0\right)$ **then**:
      i. Increase the counter $k = k + 1$.
      ii. Accept the proposed sample $\tilde{\boldsymbol{\theta}}$ as a posterior sample, i.e.:
      set $\boldsymbol{\theta}_{(k)} = \tilde{\boldsymbol{\theta}}$.
   (c) Increase the counter $n = n + 1$.
3. Estimate p_Ω as
$$p_\Omega \approx \hat{p}_\Omega = \frac{K - 1}{n - 1} \tag{10}$$
4. Evaluate the evidence $c_{\mathrm{E}} = p_\Omega \cdot c^{-1}$

On average, the algorithm requires K/p_Ω samples from the prior distribution to generate K samples from the posterior distribution. The principle of the *rejection sampling* algorithm is illustrated in Fig. 1. Note that *Algorithm (1)* is similar to a Monte Carlo simulation for solving the structural reliability problem.

3. BUS with Subset Simulation

3.1. Formulation of the limit-state function

The standard limit-state function of the BUS problem is given in Eq. (4) and Eq. (9) for the original parameter space $\boldsymbol{\Theta} \times \Pi$ and the standard Normal space $\mathbf{U} \in \mathbb{R}^{M+1}$, respectively. However, the particular format of the limit-state function of the BUS

problem is not uniquely defined. Any limit-state function that has the same probability of failure p_Ω and thus the same limit-state surface (the surface where the limit-state function equals *zero*) as $g(\boldsymbol{\theta}, \pi)$ for a given c is a valid limit-state function for the BUS problem.

For rejection sampling, the performance of the method does not depend on the particular choice of the limit-state function, because the method checks only if a sample is inside or outside of the failure domain. However, for BUS with SuS, the formulation of the limit-state function has an impact. This is related to the fact that SuS introduces intermediate failure events. These intermediate failure events are defined as $g(\boldsymbol{\theta}, \pi) \leq h_i$, where h_i is a positive constant (see Section 3.2). The particular shapes of the intermediate failure levels depend on the selected limit-state function. Loosely speaking, a smooth transition of the intermediate failure levels has a positive influence on the performance of SuS.

From a numerical point of view, the limit-state function defined in Eq. (4) and Eq. (9) is not optimal, because samples with small values of π are preferred over samples with large values of π in the initial levels of SuS (especially if prior realizations of the likelihood are small compared to c^{-1}). An alternative representation of the limit-state function that has a more appropriate shape is:

$$g_1(\boldsymbol{\theta}, \pi) = \ln(\pi) - \ln(c \cdot L(\boldsymbol{\theta}|\mathbf{d})) \tag{11}$$

where $\ln(\cdot)$ denotes the natural logarithm. The limit-state function in Eq. (11) was proposed in [21]. $g_1(\boldsymbol{\theta}, \pi)$ is obtained by applying the natural logarithm to each of the terms in Eq. (4). By comparing Eq. (11) with Eq. (4) it is obvious that both functions have the same failure domain. For enhanced numerical stability, it is usually of advantage to work with the log-transform of the likelihood, $\ln L(\boldsymbol{\theta}|\mathbf{d})$, instead of using the likelihood directly. Eq. (11) can then be expressed as:

$$g_1(\boldsymbol{\theta}, \pi) = \ln(\pi) + \ell - \ln L(\boldsymbol{\theta}|\mathbf{d}) \tag{12}$$

where $\ell = \ln(c^{-1})$. Based on Eq. (11), the intermediate failure domains can be stated as:

$$
\begin{aligned}
Z_i &= \left\{ \theta \in \mathbb{R}^M \middle| \ln(\pi) - \ln(c \cdot L(\theta|\mathbf{d})) \leq h_i \right\} \\
&= \left\{ \theta \in \mathbb{R}^M \middle| \ln(\pi) \leq \ln(c \cdot L(\theta|\mathbf{d})) + h_i \right\} \\
&= \left\{ \theta \in \mathbb{R}^M \middle| \pi \leq c \cdot L(\theta|\mathbf{d}) \cdot \exp(h_i) \right\}
\end{aligned} \tag{13}
$$

The transition of the intermediate failure levels is illustrated in Fig. 2 for the limit-state functions $g(\theta, \pi)$ and $g_1(\theta, \pi)$. Limit-state function $g_1(\theta, \pi)$ clearly is more appropriate than $g(\theta, \pi)$, because the intermediate failure domains obtained with $g_1(\theta, \pi)$ converge smoothly to the final failure domain Ω. Another viable representation of the limit-state function that ensures a smooth transition of the intermediate failure domains is [16]:

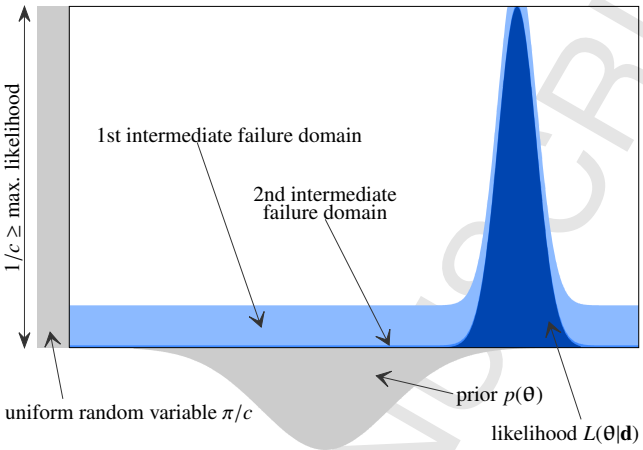
$$
g_n(\theta, \pi) = \Phi^{-1}(\pi) - \Phi^{-1}(c \cdot L(\theta|\mathbf{d})) \tag{14}
$$

where $\Phi^{-1}(\cdot)$ is the inverse of the CDF of the standard Normal distribution. In this contribution, we exclusively use limit-state function $g_1(\theta, \pi)$ as defined in Eq. (12), because it has particular advantages if the scaling parameter of *BUS-SuS* is learned adaptively (see Section 4).

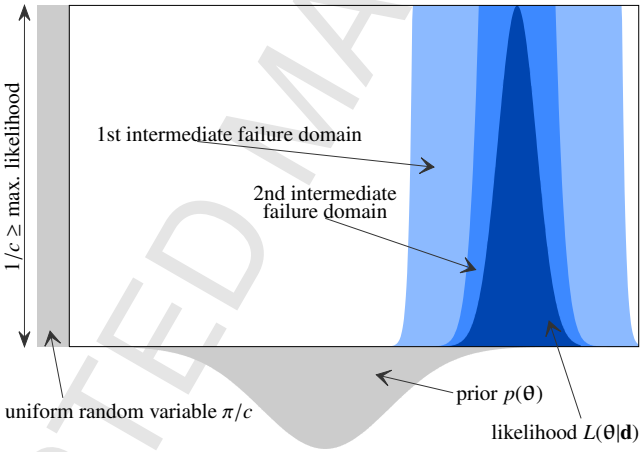
3.2. Standard Subset Simulation and BUS

SuS was proposed by Au and Beck in [20] and is an adaptive Monte Carlo method that is efficient for estimating small probabilities in high dimensional problems. SuS expresses the domain Ω as the intersection of m intermediate nested domains Z_i, where $Z_0 \supset Z_1 \supset \ldots \supset Z_m = \Omega$. The domains Z_i are defined as the sets $\{g_1(\theta, \pi) \leq h_i\}$, where h_i are threshold levels defined as $h_0 = \infty > h_1 > \ldots > h_m = 0$. Note that the intermediate failure domains Z_i are defined according to Eq. (13). We use Z_i interchangeably for domain and event. Samples conditional on Z_i are denoted $[\theta_{(i,k)}, \pi_{(i,k)}]$, for $k \in \{1, \ldots, K\}$.

By expressing Ω as the intersection of m intermediate nested domains, the small probability of failure p_Ω is equivalently expressed as the product $p_\Omega = \prod_{i=1}^{m} p_i$ of larger conditional probabilities $p_i = \Pr([\theta_{(i-1,k)}, \pi_{(i-1,k)}] \in Z_i)$. These larger conditional probabilities p_i can be estimated with Monte Carlo based techniques more efficiently



(a) BUS-limit-state function $g(\theta, \pi) = \pi - L(\theta|\mathbf{d})/c^{-1}$ defined in Eq. (4)



(b) BUS-limit-state function $g_1(\theta, \pi) = \ln(\pi) + \ell - \ln L(\theta|\mathbf{d})$ defined in Eq. (12).

Figure 2: The shape of the intermediate failure domains is shown in *(a)* and *(b)* if the BUS-limit-state function is defined according to Eq. (4) and Eq. (12), respectively. The intermediate failure domains in *(a)* do not exhibit a smooth transition to the final failure domain Ω. Therefore, this particular formulation of the limit-state function should not be used in combination with *BUS-SuS*. Instead, we recommend to use limit-state function Eq. (12) depicted in *(b)*. The intermediate failure domains obtained with this limit-state function exhibit a smooth transition to the final failure domain Ω.

The above plots were generated assuming a standard Normal prior and a Normal likelihood with mean 3 and 10% coefficient of variation. The intermediate failure domains are selected such that the conditional probability in SuS is 10%. Note that the convergence of the proposed *aBUS* algorithm is equivalent to the one of *BUS-SuS*. Both methods differ only in the shape of the final failure domain if $c^{-1} < L_{\max}$. Observing a c^{-1} not close to $L_{\max}$ for this simple example and a reasonable sample size is, however, unlikely.

than the direct estimation of p_Ω. The threshold levels h_i are typically selected such that the corresponding conditional probability p_i becomes p_t (usually chosen as $p_t = 10\%$) on average [20, 31]. The basic SuS algorithm is:

Algorithm (2): Subset Simulation for BUS

As input the algorithm requires:

- K, the total number of samples to draw from the posterior distribution.
- p_t, the probability of the intermediate subsets. p_t needs to be selected such that $p_t \cdot K$ is an integer number.
- c, selected such that $c^{-1} \geq L_{\max}$.

The algorithm evaluates the evidence c_E and returns K uniformly weighted but dependent posterior samples $\Theta_{(k)}$ with $k = 1, \ldots, K$.

1. Draw K samples $[\Theta_{(0,k)}, \pi_{(0,k)}]$, with $k = 1, \ldots, K$, from the prior distribution.
2. Initialize $i = 0$ and $h_0 = \infty$.
3. **while** $(h_i > 0)$ **do**:
   (a) Increase counter i by *one*: $i = i + 1$.
   (b) Select the threshold level h_i:
      i. Sort the K samples $\{[\Theta_{(i-1,k)}, \pi_{(i-1,k)}]\}_{k=1}^{K}$ with respect to the value of $g_1(\Theta_{(i-1,k)}, \pi_{(i-1,k)})$ in ascending order.
      ii. Set $h_i = \frac{g_1(\Theta_{(i-1,p_t \cdot K)}) + g_1(\Theta_{(i-1,p_t \cdot K+1)})}{2}$; i.e., set h_i as the p_t-percentile of the ordered set.
      iii. Select n as the number of samples in $\{[\Theta_{(i-1,k)}, \pi_{(i-1,k)}]\}_{k=1}^{K}$ with $g_1(\Theta_{(i-1,k)}, \pi_{(i-1,k)}) \leq \max(h_i, 0)$.
      iv. **if** $(h_i < 0)$ **then**: Set $h_i = 0$, and $p_i = \frac{n}{K}$.
      else: Set $p_i = p_t$.
   (c) Generate samples conditional on domain Z_i:
      i. Randomize the ordering of the samples in the set $\{[\Theta_{(i-1,k)}, \pi_{(i-1,k)}]\}_{k=1}^{n}$.

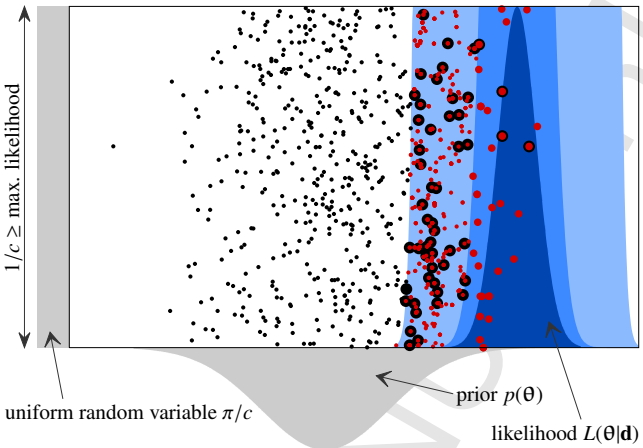


Figure 3: Illustration of the principle of the *BUS-SuS* algorithm – *Algorithm (2)*. The intermediate failure domains obtained with limit-state function $g_1(\theta, \pi)$ defined in Eq. (12) are highlighted. The innermost region is the domain Ω defined in Eq. (3); samples within this region follow the posterior distribution. The black samples are the initial samples from the prior distribution ($K = 500$ samples were used per Subset level). The large black dots indicate the $10\% \cdot K$ samples that are located in the first intermediate failure domain Z_1. These samples are used as seed values to generate samples in Z_1 by means of MCMC. The generated samples in Z_1 are highlighted in red. Note that only the *black* samples are independent; the *red* samples are dependent, because they are obtained by means of MCMC. The large red dots indicate the $10\% \cdot K$ samples that are located in the second intermediate failure domain Z_2. The number of Subset levels in this example is $m = 3$.

ii. Generate the samples $[\theta_{(i,k)}, \pi_{(i,k)}]$ by means of n Markov chains. The n samples $\{[\theta_{(i-1,k)}, \pi_{(i-1,k)}]\}_{k=1}^{n}$ are already within Z_i and are used as seeds for the n Markov chains. The length of each Markov chain is K/n. Thus, the total number of MCMC samples generated in one level is $K - n$.

4. Set $m = i$
5. Estimate $p_\Omega = \prod_{i=1}^{m} p_i$
6. Evaluate the evidence $c_\mathrm{E} = p_\Omega \cdot c^{-1}$

Note that we generate samples by means of MCMC in Step 3c(ii) of *Algorithm (2)* in the underlying *standard Normal* space (the procedure is explained in Section 3.3).

Thus, the standard Normal transform $\mathbf{u}$ of each generated sample $[\theta, \pi]$ is ideally stored as well. For the sake of convenience, this is not explicitly explained in *Algorithm (2)*.

The principle behind *Algorithm (2)* is exemplified in Fig. 3. Note that Step 3c(i) in *Algorithm (2)* is not a standard step in SuS. This step is introduced to tune the spread of the MCMC proposal distribution during the MCMC sampling – which becomes relevant in *Algorithm (4)* later. Without this step, *Algorithm (4)* would possibly introduce a bias.

In *Algorithm (2)*, the employed number of samples per level is equivalent to the number K of samples in the final level of SuS. In general, the number of samples in the final level of *BUS-SuS* can be chosen larger than the number of samples in the intermediate levels, but we do not consider this option here.

3.3. MCMC within BUS-SuS

In Step 3c(ii) of *Algorithm (2)*, MCMC sampling is applied to generate samples in domain Z_i. The n samples that are already located in Z_i are used as seed values to start n Markov chains. As the seeds of the Markov chains are already samples from the target distribution of the chains, the chains do not require a burn-in, compare [20, 31]; i.e., the chains produce samples from the target distribution right from the beginning. This property is referred to as *perfect sampling* in the context of MCMC.

The special format of the reliability problem allows the application of tailored MCMC algorithms in SuS [32]. In particular, the performance of the MCMC algorithms used in SuS does not depend on the dimensionality of the problem. A component-wise variant of the Metropolis-Hastings algorithm proposed in [20] is often used in combination with SuS. Another algorithm called *conditional sampling in U-space* that stands out due to its simplicity was proposed in [32]. The latter is applied in this contribution and presented in *Algorithm (3)*.

Algorithm (3): MCMC algorithm for SuS [32]

This algorithm works in the underlying *standard Normal space* (recall Section 2.4). Let $\mathbf{u}_{(j)}$ be the jth state of a Markov chain; remember that $\pi_{(j)} = \Phi\left((\mathbf{u}_{(j)})_{M+1}\right)$ and

$\theta_{(j)} = \mathrm{T}^*\left((\mathbf{u}_{(j)})_{1:M}\right)$. As input this algorithm requires:

- $\mathbf{u}_{(j)}$, the current state of the Markov chain (that is, a sample from the standard Normal distribution conditional on Z_i).
- i, the current level of SuS (see *Algorithm (2)*).
- $s_q \in [0, 1]$, is a parameter that controls the spread of the proposal distribution. If $s_q = 0$ the spread is *zero* and the algorithm returns the old sample. If $s_q = 1$, the spread is maximal and the algorithm returns a sample that does not depend on $\mathbf{u}_{(j)}$.

The algorithm returns the next state of the Markov chain $\mathbf{u}_{(j+1)}$ and $[\theta_{(j+1)}, \pi_{(j+1)}]$ in standard Normal space and original parameter space, respectively.

1. Propose next sample $[\tilde{\theta}, \tilde{\pi}]$ based on current $\mathbf{u}_{(j)}$:
   (a) Get the candidate sample $\tilde{\mathbf{u}}$ in standard Normal space **for each** $k = 1, \ldots, M+1$ **do**:
      i. Draw $\tilde{u}_k$ as a sample from a Normal distribution that has mean $\sqrt{1 - s_q^2} \cdot (\mathbf{u}_{(j)})_k$ and standard deviation s_q.
   (b) Set $\tilde{\theta} = \mathrm{T}^*(\tilde{\mathbf{u}}_{1:M})$ and $\tilde{\pi} = \Phi(\tilde{u}_{M+1})$.
2. Evaluate $g_1(\tilde{\theta}, \tilde{\pi})$ and check if the proposed sample $[\tilde{\theta}, \tilde{\pi}]$ is in Z_i; i.e., if $g_1(\tilde{\theta}, \tilde{\pi}) \leq h_i$.
3. **if** $\left([\tilde{\theta}, \tilde{\pi}] \in Z_i\right)$ **then** accept the sample:
   (a) Set $[\theta_{(j+1)}, \pi_{(j+1)}] = [\tilde{\theta}, \tilde{\pi}]$.
   (b) Set $\mathbf{u}_{(j+1)} = \tilde{\mathbf{u}}$.

   else reject the sample:
   (c) Set $[\theta_{(j+1)}, \pi_{(j+1)}] = [\theta_{(j)}, \pi_{(j)}]$.
   (d) Set $\mathbf{u}_{(j+1)} = \mathbf{u}_{(j)}$.

A proper selection of the parameter s_q that controls the spread of the MCMC proposal distribution is crucial. On the one hand, if s_q is selected too large, few proposed samples are accepted and the dependence of the samples in the chain is large. On the

other hand, if s_q is too small, many proposed samples are accepted, the proposed samples are in the vicinity of the previous sample and thus the dependence of the samples in the chain is also large. An adaptive strategy to optimize the spread of the MCMC proposal during the simulation is discussed in Section 3.4.

In *Algorithm (3)*, the same s_q is used for each of the $M + 1$ components. In general, the performance of *Algorithm (3)* can be improved by using different values for the individual components. However, this requires advanced information about the shape of the likelihood function – which is usually not available beforehand. Note that second-order statistics of the seeds of the Markov chains can be used to select individual values of s_q for the different components of the parameter vector [32].

3.4. Optimizing the spread of the MCMC proposal on the fly

An optimal value of s_q minimizes the dependency of the samples in the chain. However, the dependency of the samples is hard to assess and optimize during a simulation. Instead of directly assessing the dependency, the expected acceptance rate α of the Metropolis-Hastings algorithm can be monitored [32, 33]. This quantity can be easily estimated by dividing the number of samples accepted in Step 3 of *Algorithm (3)* by the total number of MCMC samples generated. The expected acceptance rate can be altered by modifying the spread of the proposal distribution. Decreasing the parameter s_q of *Algorithm (3)* decreases the spread and increases the average acceptance rate α, and vice versa.

There is an implicit connection between the dependency of the samples in the chain and the expected acceptance rate α of the Metropolis-Hastings algorithm [32, 33]. For a given problem, a value of α can be found that minimizes the dependency of the generated samples. For MCMC sampling in SuS, Papaioannou et al. [32] suggested to use a target acceptance rate α_{opt} of 0.44, which leads to near-optimal results for a range of representative examples.

The average acceptance rate of *Algorithm (3)* can be easily adapted to α_{opt} by the following strategy[32]:

Algorithm (4): Tune the spread of the MCMC proposal on the fly

The algorithm is executed after each completed Markov chain in SuS; i.e., before a new Markov chain with a new seed is started. As input this algorithm requires:

- α_{opt}, the target acceptance rate.
- s_q, the current spread of the proposal distribution.
- N_α, the number of MCMC samples after which the spread s_q is modified; e.g., $N_\alpha = 90$. The parameter n_α keeps track of how many MCMC samples were generated since the last time the spread s_q was modified. Initially, n_α is set to *zero*.
- For the first level of SuS, an initial value for parameter s_q is picked; e.g., 0.8.
- Before each subset level, the parameter N_{adpt} is set to *one* and n_α is set to *zero*.

The algorithm modifies the spread s_q of the proposal distribution.

1. Append the number of MCMC samples generated in the last Markov chain to n_α.
2. **If** $(n_\alpha \geq N_\alpha)$ **then**:
   (a) Estimate the average acceptance rate α of the past n_α samples.
   (b) Compute coefficient $c_q = \frac{\alpha - \alpha_{\text{opt}}}{\sqrt{N_{\text{adpt}}}}$.
   (c) Modify spread s_q as: $s_q = \exp(c_q) \cdot s_q$.
   (d) Increase N_{adpt} by *one*: $N_{\text{adpt}} = N_{\text{adpt}} + 1$
   (e) Set $n_\alpha = 0$.

The parameter N_{adpt} ensures that the coefficient c_q asymptotically approaches *zero* at a certain level of SuS. It is crucial that the ordering of the seeds of the Markov chain is randomized before the MCMC sampling is started (Step 3(c)i in *Algorithm (2)*) – otherwise the result of SuS is possibly biased.

4. aBUS – adaptive BUS-SuS

4.1. Introduction

Contrary to the standard algorithm presented in Section 3.2, we propose an algorithm that does not require the constant c as input. It is based on the BUS variant

originally proposed in [22] that adaptively learns the constant c. We extend the idea from [22] to improve the efficiency of the method. Due to the particular formulation of the limit-state function of the BUS problem given in Eq. (12), the method can be considerably simplified with respect to the approach presented in [22], as it requires only minimal modifications of the original *BUS-SuS* algorithm. The proposed method is termed *aBUS*, because c is selected in an adaptive manner.

For the limit-state function introduced in Eq. (11), the ith intermediate failure domain Z_i can be expressed as the set of all Θ and π for which the following inequality holds (see Eq. (13)):

$$\pi \leq c \cdot L(\Theta|\mathbf{d}) \cdot \exp(h_i) \tag{15}$$

The particular advantage of this limit-state formulation is as follows. The inequality in Eq. (15) can be equivalently stated as:

$$\pi \leq \frac{c \cdot L(\Theta|\mathbf{d}) \cdot \exp(h_i + \delta)}{\exp(\delta)} \tag{16}$$

where $\delta \in \mathbb{R}$ can be an arbitrary scalar value. From Eq. (16) it follows that the intermediate failure domain Z_i associated with scaling constant c and threshold level h_i can be equivalently expressed by scaling constant c^* and threshold level h^* if h^* is chosen as:

$$h^* = h_i + \ln\left(\frac{c}{c^*}\right) = h_i - \ell + \ell^* \tag{17}$$

where $\ell = -\ln(c)$ and $\ell^* = -\ln(c^*)$. Consequently, if one adjusts the current threshold value of SuS from h_i to h^* after changing the value of either c or ℓ to c^* or ℓ^*, the change does not affect the distribution of the current samples (i.e., the samples that are in domain Z_i), since Eq. (15) can be equivalently expressed as:

$$\pi \leq c^* \cdot L(\Theta|\mathbf{d}) \cdot \exp(h^*) \tag{18}$$

4.2. Proposed modifications to the basic BUS-SuS algorithm

The first step of the adaptive algorithm is the same as the one of standard *BUS-SuS*; it consists in drawing K samples from the prior distribution. The likelihood of each sample is evaluated and stored. However, before the value of the first threshold level h_i can be selected, a value has to be assigned to the BUS scaling constant c: The constant

c^{-1} is set equal to the value of the largest likelihood within the generated set of samples. Thereafter, each iteration is performed in accordance with *BUS-SuS*. That is, the value of each intermediate threshold is selected based on the limit-state function realizations, and then MCMC sampling is performed to generate samples conditional on the current intermediate failure domain. At the end of a subset level, it is checked whether a likelihood larger than the current value of c^{-1} was calculated. If so, the current value of c^{-1} is adapted such that it matches the largest likelihood observed and the value of h_i is modified according to Eq. (17). Note that c^{-1}, and thus the threshold h_i, can only increase. The iteration over the subset levels is performed until the current threshold value h_i is *zero* at the end of a subset level. The resulting intermediate failure domains are clearly nested – which is a prerequisite for the application of SuS. The evidence is estimated based on the last value of c at the end of SuS according to Eq. (6); i.e. with c^{-1} equal to the value of the largest likelihood observed during the simulation.

The general structure of the proposed algorithm is given in the following – changes compared to *Algorithm (2)* are highlighted in blue.

Algorithm (5): aBUS – adaptive BUS-SuS

As input the algorithm requires:

- K, the total number of samples to draw from the posterior distribution.
- p_t, the probability of the intermediate subsets. p_t needs to be selected such that $p_\mathrm{t} \cdot K$ is an integer number.

The algorithm evaluates the evidence c_E and returns K unweighted but dependent posterior samples $\boldsymbol{\Theta}_{(k)}$ with $k = 1, \ldots, K$.

1. Draw K samples $[\boldsymbol{\Theta}_{(0,k)}, \pi_{(0,k)}]$, with $k = 1, \ldots, K$, from the prior distribution.
2. Initialize $i = 0$ and $h_0 = \infty$.
3. Set $\ell = \max\left(\left\{\ln L\left(\boldsymbol{\Theta}_{(0,k)} | \mathbf{d}\right)\right\}_{k=1}^{K}\right)$, where ℓ is defined as in Eq. (12).
4. **while** $(h_i > 0)$ **do**:
   (a) Increase counter i by *one*: $i = i + 1$.
   (b) Select the threshold level h_i:

i. Sort the K samples $\{[\Theta_{(i-1,k)}, \pi_{(i-1,k)}]\}_{k=1}^{K}$ with respect to the value of $g_1(\Theta_{(i-1,k)}, \pi_{(i-1,k)})$ in ascending order.
Note that $g_1(\cdot,\cdot)$ as defined in Eq. (12) is used.

ii. Set $h_i = \frac{g_1(\Theta_{(i-1,p_t\cdot K)}) + g_1(\Theta_{(i-1,p_t\cdot K+1)})}{2}$; i.e., set h_i as the p_t-percentile of the ordered set.

iii. Select n as the number of samples in $\{[\Theta_{(i-1,k)}, \pi_{(i-1,k)}]\}_{k=1}^{K}$ with $g_1(\Theta_{(i-1,k)}, \pi_{(i-1,k)}) \leq \max(h_i, 0)$.

iv. **if** $(h_i < 0)$ **then**: Set $h_i = 0$, and $p_i = \frac{n}{K}$.
else: Set $p_i = p_t$.

(c) Generate samples conditional on domain Z_i:

i. Randomize the ordering of the samples in the set $\{[\Theta_{(i-1,k)}, \pi_{(i-1,k)}]\}_{k=1}^{n}$; i.e., thereafter, the n samples are no longer ordered.

ii. Generate the samples $[\Theta_{(i,k)}, \pi_{(i,k)}]$ by means of n Markov chains (e.g., using *Algorithm (3)* and *Algorithm (4)*). The n samples $\{[\Theta_{(i-1,k)}, \pi_{(i-1,k)}]\}_{k=1}^{n}$ are already within Z_i and are used as seeds for the n Markov chains. The length of each Markov chain is K/n. Thus, the total number of MCMC samples generated in one level is $K - n$.

(d) Update the value of the scaling constant:

i. Set $\ell_{\text{new}} = \max\left(\ell, \left\{\ln L\left(\Theta_{(i,k)}\middle|\mathbf{d}\right)\right\}_{k=1}^{K}\right)$.

ii. Modify $h_i = h_i - \ell + \ell_{\text{new}}$.

iii. Set $\ell = \ell_{\text{new}}$.

(e) Decrease dependence of the K samples:

For $(k = 1, \ldots, K)$ **do**:

i. Draw $\tilde{\pi}$ as a sample from a uniform distribution with support $[0, \min(1, \exp(\ln L(\Theta_{(i,k)}|\mathbf{d}) - \ell + h_i))]$.

ii. Set $[\Theta_{(i,k)}, \pi_{(i,k)}] = [\Theta_{(i,k)}, \tilde{\pi}]$

5. Set $m = i$
6. Estimate $p_\Omega = \prod_{i=1}^{m} p_i$
7. Evaluate the evidence $c_E = p_\Omega \cdot \exp(\ell)$

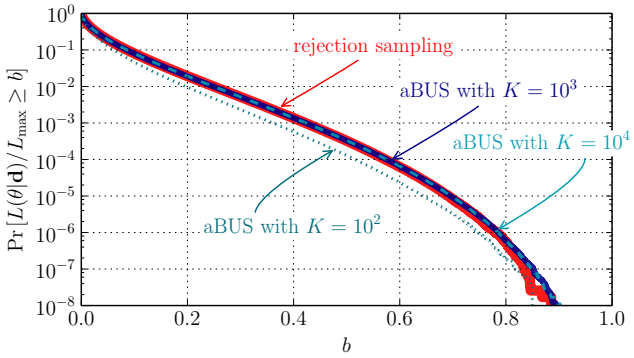


Figure 4: The posterior probability $\Pr[L(\Theta|\mathbf{d})/L_{\max} \geq b]$ is plotted for different values of b (for *Example 2*). Results are shown for posterior samples obtained by means of repeated runs of *aBUS* with $K = 10^2$, $K = 10^3$ and $K = 10^4$, where K denotes the number of samples used in each level of SuS. The reference solution is evaluated numerically by means of $6.5 \cdot 10^7$ statistically independent posterior samples obtained with rejection sampling.

In a conventional reliability problem, one typically has limited knowledge about the shape of the failure domain. Contrary to that, in reliability problems that stem from BUS, one knows that for a fixed Θ with associated likelihood $L(\Theta|\mathbf{d})$, every value of $\pi \leq cL(\Theta|\mathbf{d})$ lies in the failure domain. π conditional on $\pi \leq cL(\Theta|\mathbf{d})$ is uniformly distributed. At the intermediate levels of SuS, a sample $[\Theta, \pi]$ is in domain Z_i if $\pi \leq cL(\Theta|\mathbf{d})\exp(h_i)$. Therefore, one can easily resample the component π of sample $[\Theta, \pi]$. This is what is done in *Step 4(e)* of *Algorithm (5)*.

Step 4(e) in *Algorithm (5)* is a re-sampling strategy for π that comes at no additional cost, because it does not involve additional evaluations of the likelihood function. Its particular appeal is that in the MCMC sampling procedure, each sample that is rejected means that an existing sample is duplicated. For example, with the target acceptance rate of $\alpha_{\text{opt}} = 0.44$ suggested in Section 3.4, on average 56% percent of all proposed samples are rejected. Thus, a considerable number of the generated samples will not be unique. *Step 4(e)* distributes the π-components of all samples with the same Θ uniformly on the interval $[0, \min(1, c \cdot L(\Theta_{(i,k)}|\mathbf{d}) \cdot \exp(h_i))]$. This step is added to decrease the dependency of the generated MCMC samples and, consequently, to possibly increase the overall performance of SuS within BUS.

4.3. Comments on the final value of c in aBUS

The final value of $c^{-1} = \exp(\ell)$ in *aBUS*, corresponds to the largest likelihood observed during the simulation. Consequently, it is $\exp(\ell) \leq L_{\max}$. Asymptotically, $\exp(\ell)$ approaches $L_{\max}$ for large K, but for finite K, $\exp(\ell)$ is very likely smaller than $L_{\max}$. However, this does not prevent *aBUS* from producing samples that follow the posterior distribution, as is explained in the following.

The final $c^{-1} = \exp(\ell)$ in *aBUS* varies; it is a stochastic quantity that is equivalent to the largest likelihood value observed during the entire simulation. Let $P_L(L(\Theta)|\mathbf{d})$ denote the CDF of likelihood values evaluated for samples of the posterior distribution. If aBUS produced independent posterior samples, the quantity c^{-1} would be a realization from a distribution that has CDF $P_{c^{-1}}(c^{-1}|\mathbf{d}) = \left(P_L\left(c^{-1}|\mathbf{d}\right)\right)^K$.

In the following, we show that the CDF $P_L(L(\Theta)|\mathbf{d})$ can be approximated well with posterior samples generated with *aBUS*. The decisive parameter in *aBUS* is the number K of samples employed in each level of SuS. For the method to generate posterior samples, K must be selected large enough such that the final K samples can propagate over the entire domain Ω. The bulk of the generated posterior samples will be in the "high probability region" of the posterior distribution – which does not necessarily mean that many samples will fall in the region that has large likelihood.

We numerically investigate the distribution of likelihood values associated with the generated posterior samples for *Example 2* that is described in detail in Section 5.2. This problem demonstrates numerically that the probability that we will observe a likelihood value larger than $0.8 \cdot L_{\max}$ in a set of 10^3 independent posterior samples is $1 \cdot 10^{-4}$ (see Table 1). Therefore, it is unlikely that the value of c^{-1} in *aBUS* is close to the theoretical $L_{\max}$ for 10^3 generated samples. The Bayesian inference problem is solved with *aBUS* for $K = 10^2$, $K = 10^3$ and $K = 10^4$ samples per subset level. The reference distribution is obtained by means of rejection sampling and $c = L_{\max}^{-1}$. The posterior probability $\Pr[L(\Theta|\mathbf{d})/L_{\max} \geq b]$ estimated with *aBUS* and rejection sampling is depicted in Fig. 4 for different values of b and K (the definition of b is according to Section 5.1). For $K = 10^2$, the resulting posterior distribution of the likelihood values deviates from the reference solution. For $K = 10^3$, the resulting posterior distribution matches the reference solution well. Therefore, even if *aBUS* selects c^{-1} considerably

smaller than $L_{\max}$, the distribution of the likelihood values associated with the generated posterior samples does not exhibit a relevant bias – provided that K is selected large enough. The performance of *aBUS* with respect to K is investigated in detail in Section 5.

Note that the average number of likelihood function calls in *aBUS* is at most as large as in standard *BUS-SuS*. This is because $c_m^{-1} \leq L_{\max}$ in *aBUS*, whereas for *BUS-SuS*, $c^{-1} \geq L_{\max}$ is required. Thus, as p_Ω is proportional to c^{-1}, the average total number of subset levels required to solve the inference problem with *aBUS* is at most as large as the one of *BUS-SuS*. In addition, if the limit-state formulation given in Eq. (11) is used, the distribution of samples produced at the intermediate levels of SuS is invariant to the selected c – compare Section 4.1 and Eq. (16) in particular. The dependency of the generated samples tends to increase with increasing level in SuS. Thus, a smaller number of subset levels is preferable in BUS-SuS with respect to the statistics of the generated posterior samples. As a consequence, the *aBUS* algorithm should be preferred over standard *BUS-SuS* even if the theoretical maximum of the likelihood function is known in advance.

5. Illustrative applications

5.1. Definitions

For the discussion of the problems, we introduce the following quantities:

- b acts as a normalized version of c^{-1}: Let $b \in (0, 1]$ be defined as $b = 1/(c \cdot L_{\max})$; i.e., $b = 1 \Leftrightarrow c^{-1} = L_{\max}$ and $b = 0 \Leftrightarrow c^{-1} = 0$. The performance of the investigated algorithm is assessed for different values of b.

- $b_{10^3,\max}$ represents the largest observed likelihood multiplied with c in a set of 10^3 independent posterior samples. Note that $b_{10^3,\max}$ is a stochastic quantity.

- $c_{\mathrm{E,ref}}$ denotes the actual value of the evidence of the investigated example. The quantity $p_{\Omega,\mathrm{ref}}$ is defined as $p_{\Omega,\mathrm{ref}} = c_{\mathrm{E,ref}}/L_{\max}$.

– $\hat{c}_{\mathrm{E},K}$ is the evidence estimated by the investigated algorithm based on K posterior samples. The bias in the estimated evidence is denoted as:

$$\mathrm{bias}\left[\hat{c}_{\mathrm{E},K}\right] = \left|\frac{\mathrm{E}\left[\hat{c}_{\mathrm{E},K}\right] - c_{\mathrm{E}}}{c_{\mathrm{E}}}\right| \tag{19}$$

The coefficient of variation of the estimated evidence is:

$$\mathrm{CoV}\left[\hat{c}_{\mathrm{E},K}\right] = \frac{\sigma\left[\hat{c}_{\mathrm{E},K}\right]}{\mathrm{E}\left[\hat{c}_{\mathrm{E},K}\right]} \tag{20}$$

– a_K and s_K denote the estimated mean and standard deviation of the first component of the parameter vector in a set of K posterior samples. Note that a_K and s_K are random variables for finite K. If the investigated algorithm produces posterior samples, we have $\mathrm{E}[a_K] = \mathrm{E}[\theta_1|\mathbf{d}]$ and $\mathrm{E}[s_K] = \sigma[\theta_1|\mathbf{d}]$. If the generated posterior samples are independent, then $\sigma[a_K] = \frac{1}{\sqrt{K}} \cdot \sigma[\theta_1|\mathbf{d}]$. For dependent samples, $\sigma[a_K]$ can be expressed as

$$\sigma[a_K] = \sqrt{\frac{1+\gamma}{K}} \cdot \sigma[\theta_1|\mathbf{d}] \tag{21}$$

where $\gamma \geq 0$ quantifies the dependency of the generated samples.

The bias in the estimated posterior mean and posterior standard deviation is defined as:

$$\mathrm{bias}\left[a_K\right] = \left|\frac{\mathrm{E}\left[a_K\right] - \mathrm{E}[\theta_1]}{\mathrm{E}[\theta_1]}\right| \tag{22}$$

$$\mathrm{bias}\left[s_K\right] = \left|\frac{\mathrm{E}\left[s_K\right] - \sigma[\theta_1]}{\sigma[\theta_1]}\right| \tag{23}$$

- N_{eff} is the number of effectively independent samples in the generated set of K posterior samples (of the first component θ_1). This quantity specifies how many truly independent posterior samples of θ_1 would give the same variance in the sample mean as $\mathrm{Var}[a_K]$ obtained by aBUS.

$$N_{\mathrm{eff}} = \left(\frac{\mathrm{E}\left[s_K\right]}{\sigma\left[a_K\right]}\right)^2 = \frac{K}{1+\gamma} \tag{24}$$

Note that N_{eff} can be interpreted as a measure for the dependency of the generated posterior samples; the smaller N_{eff} the stronger the dependency. The definition

Table 1: Reference solution of the investigated examples.

	Example 1a	**Example 1b**	**Example 2**	**Example 3**
$c_{\mathrm{E,ref}}$	$6.16 \cdot 10^{-3}$	$2.36 \cdot 10^{-6}$	$1.00 \cdot 10^{-6}$	$1.52 \cdot 10^{-3}$
$L_{\max}$	1.33	1.99	$7.47 \cdot 10^{-3}$	1.00
$p_{\Omega,\mathrm{ref}}$	$4.63 \cdot 10^{-3}$	$1.18 \cdot 10^{-6}$	$1.34 \cdot 10^{-3}$	$1.52 \cdot 10^{-3}$
$\mathrm{E}[\theta_1]$	2.75	4.81	0.34	1.12
$\sigma[\theta_1]$	0.287	0.196	0.51	0.66
$\mathrm{E}[b_{10^3,\max}]$	1	1	0.46	0.999
$\Pr\left[b_{10^3,\max} > 0.8\right]$	1	1	$1 \cdot 10^{-4}$	1
$\Pr\left[b_{10^3,\max} < 0.99\right]$	10^{-37}	10^{-32}	1	$1 \cdot 10^{-4}$
$\Pr\left[b_{10^3,\max} < 0.999\right]$	$5 \cdot 10^{-12}$	$1 \cdot 10^{-10}$	1	0.38

of N_{eff} is not unique. It can be defined for any component θ_i of the parameter vector $\boldsymbol{\theta}$, or for any function of $\boldsymbol{\theta}$.

The expectations in Eqs. (19), (20), (22) and (23) are estimated by their corresponding sample averages through performing multiple runs of aBUS.

5.2. Investigated examples

– *Example 1a:* A one dimensional problem with a standard Normal prior. The uncertain parameter is denoted by θ. The likelihood of θ is a Normal distribution that has mean $\mu_l = 3$ and standard deviation $\sigma_l = 0.3$. This problem has an analytical solution: the posterior distribution is Normal with mean and standard deviation of $\mu_l / \left(\sigma_l^2 + 1\right) = 2.75$ and $1/\sqrt{1+\sigma_l^{-2}} = 0.287$, respectively. The maximum of the likelihood is $L_{\max} = 1/\left(\sigma_l \sqrt{2\pi}\right) = 1.33$. The evidence associated with this example is $c_{\mathrm{E,ref}} = \varphi\left(\mu_l / \sqrt{1+\sigma_l^2}\right) / \sqrt{1+\sigma_l^2} = 6.16 \cdot 10^{-3}$, where $\varphi(\cdot)$ is the PDF of the standard Normal distribution. Consequently, $p_{\Omega,\mathrm{ref}}$ of the rejection sampling algorithm is $4.63 \cdot 10^{-3}$, if $c = 1/L_{\max}$.

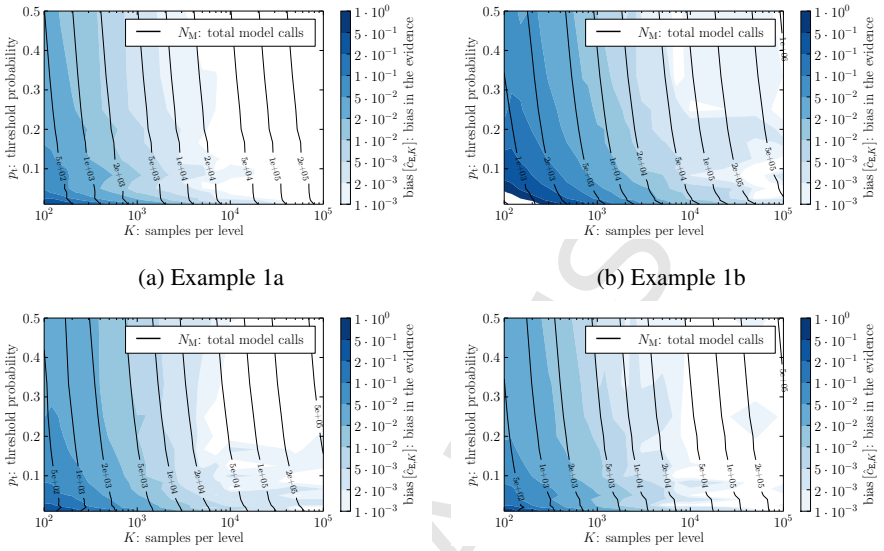
– *Example 1b:* The formulation of this problem is equivalent to *Example 1a*, with the only difference being that the likelihood function for θ has mean $\mu_l = 5$ and standard deviation $\sigma_l = 0.2$. The posterior mean and standard deviation are 4.81

and 0.196, respectively. The evidence for this problem is $c_{\mathrm{E,ref}} = 2.36 \cdot 10^{-6}$. Consequently, $p_{\Omega,\mathrm{ref}}$ of the rejection sampling algorithm is $1.18 \cdot 10^{-6}$, if $c = 1/L_{\max}$ and $L_{\max} = 1.99$.

– *Example 2:* A 12-dimensional problem with prior $\prod_{i=1}^{12} \varphi(\theta_i)$, where $\varphi(\cdot)$ denotes the standard Normal PDF and θ_i is the ith component of the 12-dimensional parameter vector $\boldsymbol{\theta}$. The likelihood function of the problem is $\prod_{i=1}^{12} \varphi\left(\frac{\theta_i - \mu_l}{\sigma_l}\right)/\sigma_l$, with $\sigma_l = 0.6$. The value μ_l is chosen such that the evidence $c_{\mathrm{E,ref}}$ becomes 10^{-6}; i.e., $\mu_l = 0.462$. The posterior mean and standard deviation of each component of $\boldsymbol{\theta}$ are 0.34 and 0.51, respectively. The theoretical maximum that the likelihood function can take is $L_{\max} = \left(0.6 \cdot \sqrt{2\pi}\right)^{-12} = 7.47 \cdot 10^{-3}$. Thus, $p_{\Omega,\mathrm{ref}}$ of the rejection sampling algorithm is $1.34 \cdot 10^{-4}$, if $c = 1/L_{\max}$.

– *Example 3:* A two-story frame structure represented as a two-degree-of-freedom shear building model is investigated. This example was originally discussed in [34]. BUS is applied in [16, 21] to solve this problem. The two stiffness coefficients k_1 (first story) and k_2 (second story) of the model are considered uncertain. The uncertainty in k_1 and k_2 is expressed as $k_1 = \theta_1 \cdot k_{\mathrm{n}}$ and $k_2 = \theta_2 \cdot k_{\mathrm{n}}$, where θ_1 and θ_2 are uncertain parameters and $k_{\mathrm{n}} = 29.7 \cdot 10^6$N/m. The prior distributions of θ_1 and θ_2 are modeled as independent log-Normal distributions with modes 1.3 and 0.8 and standard deviation 1.0. The lumped story masses m_1 (first story) and m_2 (second story) are considered deterministic and have masses $m_1 = 16.5 \cdot 10^3$kg and $m_2 = 16.1 \cdot 10^3$kg. The influence of damping is neglected. Bayesian updating is performed based on the measured first two eigen-frequencies of the system: $\tilde{f}_1 = 3.13$Hz and $\tilde{f}_2 = 9.83$Hz. The likelihood of the problem is expressed as $L(\boldsymbol{\theta}) = \exp\left(-0.5 \cdot J(\boldsymbol{\theta})/\sigma_\varepsilon^2\right)$, where $\sigma_\varepsilon = 1/16$ and $J(\boldsymbol{\theta}) = \sum_{j=1}^{2} \lambda_j^2 \left(\frac{f_j^2(\boldsymbol{\theta})}{\tilde{f}_j^2} - 1\right)^2$ with $\lambda_1 = \lambda_2 = 1$ and $f_j(\boldsymbol{\theta})$ as the jth eigen-frequency predicted by the model. The posterior distribution of this problem is bimodal [16, 34]. The reference evidence is: $c_{\mathrm{E,ref}} = p_{\Omega,\mathrm{ref}} = 1.52 \cdot 10^{-3}$ (since $L_{\max} = 1$). Moreover, $\mathrm{E}[k_1|\mathbf{d}] = 1.12$ and $\sigma[k_1|\mathbf{d}] = 0.66$.

The reference solutions of the presented examples are summarized in Table 1. In addition to the quantities $c_{\mathrm{E,ref}}$, $L_{\max}$, $p_{\Omega,\mathrm{ref}}$, $\mathrm{E}[\theta_1|\mathbf{d}]$ and $\sigma[\theta_1|\mathbf{d}]$, some statistics of quan-



(a) Example 1a (b) Example 1b

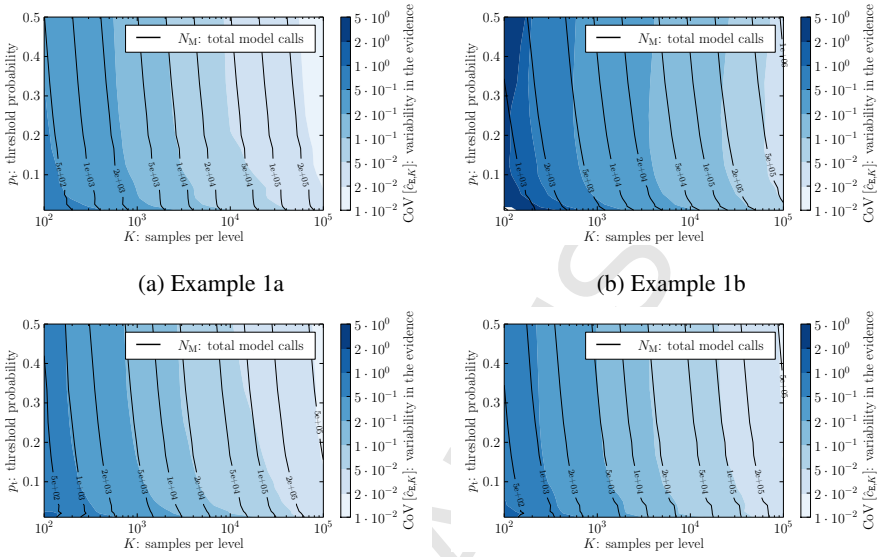
(c) Example 2 (d) Example 3

Figure 5: Bias in the evidence estimated with *aBUS* for different p_t and K.

tity $b_{10^3,\max}$ are listed in the last four rows. The statistics for $b_{10^3,\max}$ can be computed explicitly for *Examples 1a*, *1b* and *2*, and are evaluated numerically by solving the updating problem $7 \cdot 10^5$ times with the rejection sampling algorithm and $K = 10^3$ in each run for *Example 3*. It is obvious that *Example 2* differs from the other problems with respect to the statistics of $b_{10^3,\max}$: For *Example 2*, $\mathrm{E}[b_{10^3,\max}] = 0.46$ and $\Pr\left[b_{10^3,\max} > 0.8\right] = 1 \cdot 10^{-4}$, whereas $\mathrm{E}[b_{10^3,\max}] \approx 1$ and $\Pr\left[b_{10^3,\max} > 0.8\right] = 1$ for all other examples. Consequently, it is unlikely that a $b_{10^3,\max}$ close to *one* will be observed in *Example 2* in a set of 10^3 posterior samples.

5.3. Performance of aBUS for different p_t and K

The performance of *aBUS* is assessed for different p_t and K. The probability of the intermediate subsets p_t is analyzed for values in $[1\%, 50\%]$. The number K of samples per level is modified between 10^2 and 10^5. Our aim is to determine which values of p_t lead to a (near-)optimal performance for the investigated examples, where optimality is measured with respect to the number N_M of total required model calls; i.e. N_M is the total number of likelihood evaluations in SuS. For the MCMC sampling in the subset



(a) Example 1a

(b) Example 1b

(c) Example 2

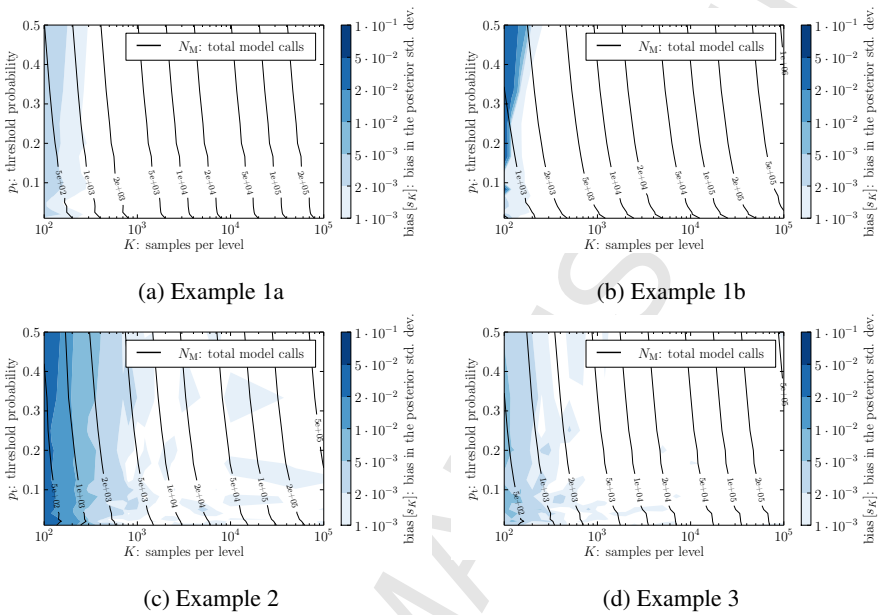
(d) Example 3

Figure 6: Coefficient of variation of the evidence estimated with *aBUS* for different p_t and K.

levels we employ *Algorithm (3)*. The spread of the proposal distribution is modified during the simulation as described in *Algorithm (4)*, with a target acceptance rate of $\alpha_{\text{opt}} = 0.44$.

Fig. 5 presents the bias in the evidence estimated with *aBUS* (by means of measure bias $[\hat{c}_{E,K}]$ introduced in Eq. (19)). The bias in the evidence decreases with an increasing number of samples per level in all investigated examples. We observe that $p_t = 10\%$ is clearly not an optimal choice. Especially for $K < 10^3$, the bias is smaller for large p_t than for small p_t. An intermediate probability p_t between 20% and 40% is a good choice for all investigated problems. Among all investigated examples, the largest bias is observed in *Example 1b*; the smallest bias is observed in *Example 1a*. This suggests that the bias in the evidence computed with *aBUS* (and probably *BUS-SuS* in general) increases with an increasing number of subset levels.

Overall, the bias in the estimated evidence of *aBUS* is negligible compared to the coefficient of variation of the estimate. The coefficient of variation CoV $[\hat{c}_{E,K}]$ in the estimated evidence is depicted in Fig. 6. The CoV $[\hat{c}_{E,K}]$ decreases with an increasing



(a) Example 1a (b) Example 1b

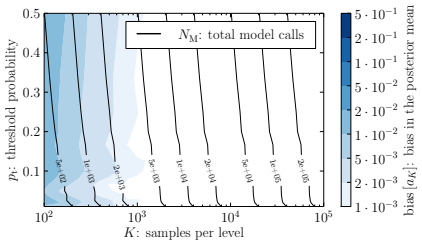
(c) Example 2 (d) Example 3

Figure 7: Bias in the mean of posterior samples generated with *aBUS* for different p_t and K.

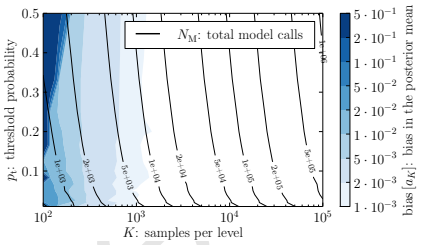
number of samples per level. For p_t between 10% and 30%, *aBUS* performs robustly with respect to a fixed number N_M of total model calls in all investigated examples.

Fig. 7 and Fig. 8 present the bias in the estimated mean and standard deviation of the posterior samples produced with *aBUS*. For $K \geq 5 \cdot 10^2$, the bias in both mean and standard deviation is smaller than 0.5% and is considered negligible.

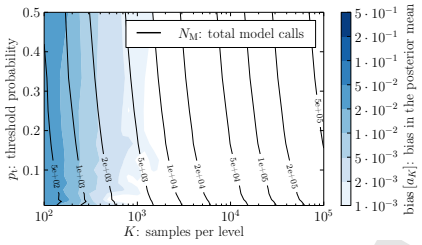
Finally, Fig. 9 summarizes the number N_{eff} of effectively independent samples in the generated set of K posterior samples. N_{eff} increases with increasing K. For a fixed number N_M of total model calls, *aBUS* exhibits the best performance for $p_t = 10\%$. However, N_{eff} is always considerably smaller than K. For $K = 10^3$ and $p_t = 10\%$ we obtain only 210, 150, 70 and 20 effectively independent posterior samples in *Example 1a*, *1b*, *2* and *3*, respectively. In particular, $N_{eff} \approx 20$ in *Example 3* is a relatively small value. The poor performance in this example can be attributed to the bimodal shape of the posterior distribution: The standard deviation of quantity a_K that governs N_{eff} (see Eq. (24)) is relatively large in this example, because it is difficult for the intermediate samples in SuS to alternate between the two modes. If the fraction of samples in the



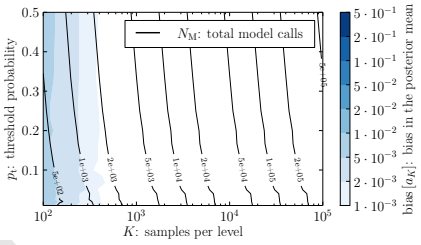
(a) Example 1a



(b) Example 1b

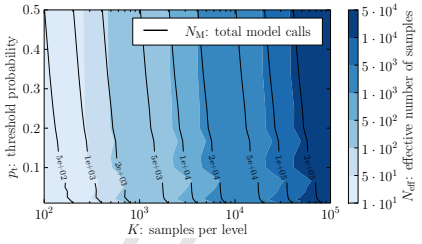


(c) Example 2

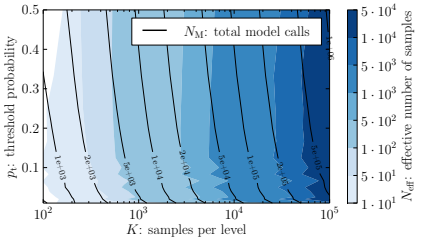


(d) Example 3

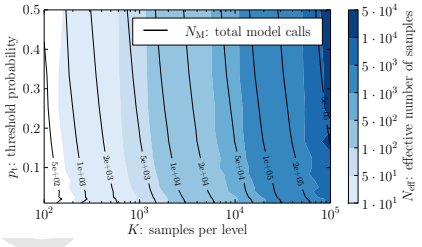
Figure 8: Bias in the standard deviation of posterior samples generated with *aBUS* for different p_t and K.



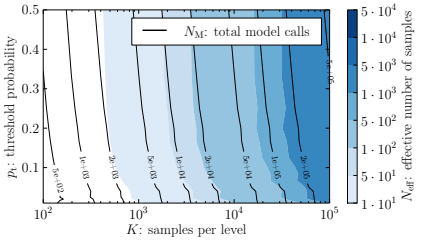
(a) Example 1a



(b) Example 1b



(c) Example 2



(d) Example 3

Figure 9: Number N_{eff} of effectively independent samples obtained with *aBUS* for different p_t and K.

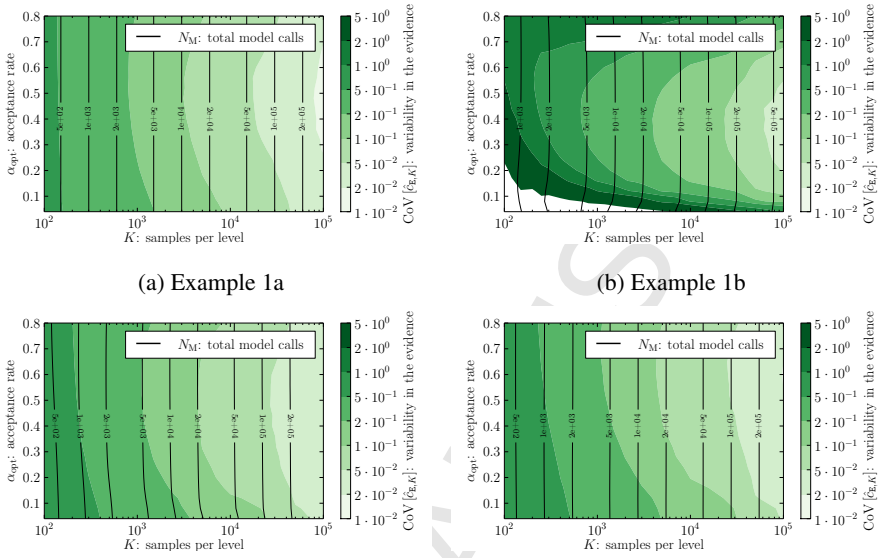
separate modes at initial subset levels is amiss, this error will likely propagate to the higher levels of SuS. However, *Example 3* demonstrates also the flexibility of *BUS-SuS* based approaches: They are able to produce posterior samples even if the target distribution is multi-modal.

To summarize the findings obtained in this section: The potential bias in the evidence estimated with *aBUS* is negligible compared to the variability of the estimate. Furthermore, the bias in the mean and standard deviation of posterior samples produced with *aBUS* is insignificant for reasonable sample sizes. For the investigated examples, $K \geq 5 \cdot 10^2$ was large enough. However, as a general rule of thumb, the authors recommend to use at least 10^3 samples per level in SuS. Therefore, the parameter p_t of SuS should be selected such that for a given number K of posterior samples, N_eff is maximized and $\mathrm{CoV}\left[\hat{c}_{\mathrm{E},K}\right]$ is minimized. Based on the investigated examples, we suggest using a value of $p_\mathrm{t} = 10\%$, which provides a reasonable compromise for the performance with respect to N_eff and the coefficient of variation of the estimated evidence.

5.4. Performance of aBUS for different α_{opt} and K

In this study, we modify the target acceptance rate α_opt of *Algorithm (4)* and the number K of samples per level: α_opt is changed between 0.04 and 0.80, and K is modified between 10^2 and 10^5. Because the bias of *aBUS* for the estimated evidence, the posterior mean and the posterior standard deviation were found to be negligible in the previous study (Section 5.3), we only investigate the performance in terms of $\mathrm{CoV}\left[\hat{c}_{\mathrm{E},K}\right]$ and N_eff. Again, we assess the performance of *aBUS* for combinations of α_opt and K that result in the same number N_M of total likelihood evaluations during SuS. In this study, the total number of required likelihood evaluations is approximately proportional to K. The probability of the intermediate subsets p_t is kept constant at 10%.

The coefficient of variation of the evidence estimated with *aBUS* is shown in Fig. 10 for different α_opt and K. For fixed N_M, a comparatively good performance is achieved for all investigated examples if α_opt is selected between 0.4 and 0.6; where *Example 1b* favors slightly smaller α_opt for large K and *Example 3* favors slightly larger α_opt. For



(a) Example 1a (b) Example 1b

(c) Example 2 (d) Example 3

Figure 10: Coefficient of variation of the evidence estimated with *aBUS* for different α_{opt} and K.

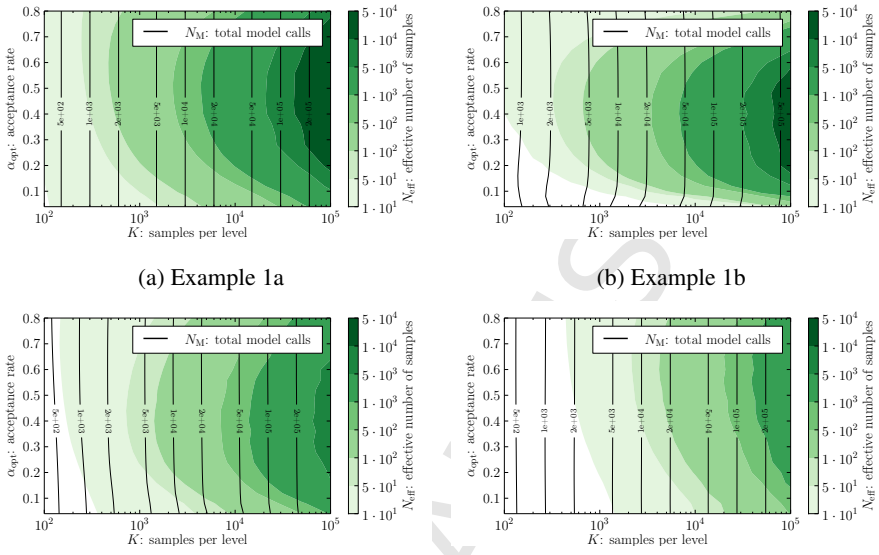
Example 1b that has the smallest p_Ω amongst all investigated examples, the dependency of CoV $[\hat{c}_{E,K}]$ on α_{opt} is more pronounced than in the other problems.

The number N_{eff} of effectively independent posterior samples in the generated set of K posterior samples is depicted in Fig. 11. For *Examples 1a*, *1b* and *2*, *aBUS* exhibits a near-optimal performance for $0.3 \leq \alpha_{\text{opt}} \leq 0.5$ (with respect to N_M fixed). For *Example 3*, a slightly better performance is achieved for larger values of α_{opt}; i.e. for α_{opt} selected around 0.6. Again, it is evident that for *Example 3*, *aBUS* produces a relatively small N_{eff}.

In summary, the choice of $\alpha_{\text{opt}} = 0.44$ proposed in [32] for SuS works reasonably well for *aBUS*. The number N_{eff} of effectively independent posterior samples in the generated set of K posterior samples depends strongly on the problem at hand.

5.5. Performance in high dimensions

The behavior of aBUS for a large number of uncertain parameters is studied by means of a numerical example, which was also used in [22]. The components $\{\theta_i : i = 1, \ldots, M\}$ of M-dimensional random vector $\boldsymbol{\theta}$ are independent and have identical



(a) Example 1a (b) Example 1b

(c) Example 2 (d) Example 3

Figure 11: Effective number N_{eff} of independent samples obtained with *aBUS* for different α_{opt} and K.

distributions. The prior PDF of $\boldsymbol{\theta}$ is $p(\boldsymbol{\theta}) = \prod_{i=1}^{M} \varphi(\theta_i)$, where $\varphi(\cdot)$ is the PDF of the univariate standard Normal distribution. The likelihood of the problem is defined as $L(\boldsymbol{\theta}|\mathbf{d}) = \frac{1}{\sigma_\varepsilon}\varphi\left(\frac{h(\boldsymbol{\theta})-\mu_\varepsilon}{\sigma_\varepsilon}\right)$, with $h(\boldsymbol{\theta}) = \frac{1}{\sqrt{M}}\sum_{i=1}^{M}\theta_i$, $\mu_\varepsilon = 4$ and $\sigma_\varepsilon = 0.2$. This problem can be solved analytically: The evidence $c_{\text{E}} = \frac{1}{\sqrt{1+(\sigma_\varepsilon)^2}}\varphi\left(\frac{\mu_\varepsilon}{\sqrt{1+(\sigma_\varepsilon)^2}}\right) = 1.785 \cdot 10^{-4}$ is independent of M. The posterior mean and standard deviation of $h(\boldsymbol{\theta})$ is $\text{E}[h(\boldsymbol{\theta})|\mathbf{d}] = \mu_\varepsilon/(\sigma_\varepsilon^2 + 1) = 3.846$ and $\sigma[h(\boldsymbol{\theta})|\mathbf{d}] = \sqrt{\sigma_\varepsilon^2/(\sigma_\varepsilon^2 + 1)} = 0.196$, respectively. In this example, a_K and s_K are defined with respect to random quantity $h(\boldsymbol{\theta})$, and so is N_{eff}. This is because the distribution of the individual components θ_i depends on the dimension M, but the distribution of the average $h(\boldsymbol{\theta})$ does not depend on M.

Results for various dimensions M are listed in Table 2. The results were obtained through repeated runs of aBUS with $K = 10^3$, $p_{\text{t}} = 10\%$ and $\alpha_{\text{opt}} = 0.44$. The obtained results do not depend on the number of uncertain parameters in the problem. Only the bias in the evidence tends to increase slightly with increasing M.

Table 2: Statistics of the aBUS solution of the high-dimensional problem introduced in Section 5.5 for a repeated number of simulation runs with $K = 10^3$, $p_t = 10\%$ and $\alpha_{opt} = 0.44$ for various dimensions M.

M	bias $[\hat{c}_{E,K}]$	CoV $[\hat{c}_{E,K}]$	bias $[a_K]$	bias $[s_K]$	N_{eff}
1	1.8%	29%	10^{-4}	10^{-3}	176
2	1.8%	29%	10^{-4}	10^{-3}	176
10	1.9%	29%	10^{-4}	10^{-3}	179
10^2	2.1%	29%	10^{-4}	10^{-3}	176
10^3	2.3%	29%	10^{-4}	10^{-3}	175
10^4	1.2%	29%	10^{-4}	10^{-3}	171
10^5	4.0%	30%	10^{-4}	10^{-3}	170

6. Concluding remarks

Bayesian Updating with Structural reliability methods (BUS) reinterprets the Bayesian updating problem as a structural reliability problem. The evidence for the model class based on the data is obtained as a by-product of BUS. The BUS-problem can be solved efficiently with Subset Simulation (SuS); a reliability method that can efficiently handle a large number of uncertain parameters and estimate very small probabilities. Thus, the combination of BUS with SuS (*BUS-SuS*) is ideally suited for inference problems with a large number of uncertain parameters to be learned. *BUS-SuS* is also suitable for inference problems where the model behind the likelihood function is computationally demanding, because *BUS-SuS* produces posterior samples with a reasonable number of total model calls. A limitation of BUS is the need to select a scaling parameter c prior to the analysis. On the one hand, c^{-1} must be larger than the maximum of the likelihood function. On the other hand, the larger c^{-1}, the less efficient is the solution of the problem.

In this paper, we propose an extended variant of *BUS-SuS*, termed *aBUS*, that does not require the choice of the scaling constant c. *aBUS* adaptively selects appropriate values for c^{-1}. The method can be easily implemented in existing software packages that support SuS, because only slight modifications to the SuS implementation are necessary. We demonstrate with numerical examples that, for a reasonably large number

K of samples per level, *aBUS* produces samples that follow the posterior distribution – even if the constant c^{-1} in *aBUS* is found smaller than the maximum of the likelihood function.

The average number of total likelihood evaluations required by *aBUS* is smaller than or equal to the one required by standard *BUS-SuS* with fixed c because *aBUS* needs at most as many subset levels as *BUS-SuS*. Moreover, for the particular limit-state function applied within aBUS in this contribution, the distribution of the samples at the intermediate levels of SuS is invariant to the selected c. This means that *aBUS* and *BUS-SuS* essentially differ in their respective stopping criteria – but not in the intermediate sampling procedure. Because *aBUS* can stop before *BUS-SuS* but not later, and given that the dependency of the produced samples in SuS increases with the required intermediate levels, we suggest using *aBUS* instead of *BUS-SuS* even if the theoretical maximum of the likelihood function is known.

For SuS, the probability of the intermediate levels needs to be specified. Our numerical investigations show that for aBUS the commonly used conditional probability of $p_t = 10\%$ for the intermediate levels is close to the optimum in many cases. Additionally, a good value of the target acceptance rate of the MCMC sampling procedure for optimizing the spread of the proposal distribution is $\alpha_{opt} = 0.44$, as originally proposed in [32] for reliability analysis.

Acknowledgments

We acknowledge the support of the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative.

References

[1] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Introducing Markov chain Monte Carlo," *Markov chain Monte Carlo in practice*, vol. 1, p. 19, 1996.

[2] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, "Bayesian Data Analysis," 2003.

[3] M. Plummer, N. Best, K. Cowles, and K. Vines, "CODA: Convergence diagnosis and output analysis for MCMC," *R news*, vol. 6, no. 1, pp. 7–11, 2006.

[4] H. Haario, E. Saksman, and J. Tamminen, "Componentwise adaptation for high dimensional MCMC," *Computational Statistics*, vol. 20, no. 2, pp. 265–273, 2005.

[5] G. Robert and R. Tweedie, "Exponential convergence of Langevin diffusions and their discrete approximation," *Bernoulli*, vol. 2, pp. 341–363, 1996.

[6] R. M. Neal, "MCMC using Hamiltonian dynamics," in *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, eds.), ch. 5, pp. 113–162, Chapman & Hall, 2011.

[7] S. H. Cheung and J. L. Beck, "Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters," *Journal of Engineering Mechanics*, vol. 135, no. 4, pp. 243–255, 2009.

[8] J. L. Beck, "Bayesian system identification based on probability logic," *Structural Control and Health Monitoring*, vol. 17, no. 7, pp. 825–847, 2010.

[9] J. L. Beck and K.-V. Yuen, "Model selection using response measurements: Bayesian probabilistic approach," *Journal of Engineering Mechanics*, 2004.

[10] D. J. MacKay, "Bayesian interpolation," *Neural computation*, vol. 4, no. 3, pp. 415–447, 1992.

[11] L. Wasserman, "Bayesian model selection and model averaging," *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 92–107, 2000.

[12] J. Beck and L. Katafygiotis, "Updating of a model and its uncertainties utilizing dynamic test data," in *Computational Stochastic Mechanics*, pp. 125–136, Springer, 1991.

[13] J. L. Beck and L. S. Katafygiotis, "Updating models and their uncertainties. I: Bayesian statistical framework," *Journal of Engineering Mechanics*, vol. 124, no. 4, pp. 455–461, 1998.

[14] D. J. MacKay, *Bayesian methods for adaptive models*. PhD thesis, California Institute of Technology, 1992.

[15] S. H. Cheung and J. L. Beck, "Calculation of posterior probabilities for Bayesian model class assessment and averaging from posterior samples based on dynamic system data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, no. 5, pp. 304–321, 2010.

[16] D. Straub and I. Papaioannou, "Bayesian updating with structural reliability methods," *Journal of Engineering Mechanics*, 2015.

[17] O. Ditlevsen and H. O. Madsen, *Structural reliability methods*. Wiley New York, 1996.

[18] R. E. Melchers, *Structural reliability analysis and prediction*. John Wiley & Son Ltd, 1999.

[19] D. Straub, "Engineering risk assessment," in *Risk – A Multidisciplinary Introduction*, pp. 333–362, Springer, 2014.

[20] S.-K. Au and J. L. Beck, "Estimation of small failure probabilities in high dimensions by subset simulation," *Probabilistic Engineering Mechanics*, vol. 16, no. 4, pp. 263–277, 2001.

[21] F. DiazDelaO, A. Garbuno-Inigo, S. Au, and I. Yoshida, "Bayesian updating and model class selection with Subset Simulation," *Computer Methods in Applied Mechanics and Engineering*, 2017.

[22] W. Betz, I. Papaioannou, and D. Straub, "Adaptive variant of the BUS approach to Bayesian updating," in *Proceedings of 9th European Conference on Structural Dynamics*, 2014.

[23] M. Chiachio, J. L. Beck, J. Chiachio, and G. Rus, "Approximate Bayesian computation by subset simulation," *SIAM Journal on Scientific Computing*, vol. 36, no. 3, pp. A1339–A1358, 2014.

[24] M. K. Vakilzadeh, Y. Huang, J. L. Beck, and T. Abrahamsson, "Approximate Bayesian Computation by Subset Simulation using hierarchical state-space models," *Mechanical Systems and Signal Processing*, 2016.

[25] D. G. Giovanis, I. Papaioannou, D. Straub, and V. Papadopoulos, "Bayesian updating with subset simulation using artificial neural networks," *Computer Methods in Applied Mechanics and Engineering*, vol. 319, pp. 124–145, 2017.

[26] I. Papaioannou, W. Betz, and D. Straub, "Bayesian model updating of a tunnel in soft soil with settlement measurements," *Geotechnical Safety and Risk IV*, p. 351, 2013.

[27] W. Betz, C. M. Mok, I. Papaioannou, and D. Straub, "Bayesian model calibration using structural reliability methods: application to the hydrological abc model," in *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management* (M. Beer, S.-K. Au, and J. W. Hall, eds.), pp. 2734–2743, ASCE, 2014.

[28] M. Hohenbichler and R. Rackwitz, "Non-normal dependent vectors in structural safety," *Journal of the Engineering Mechanics Division*, vol. 107, no. 6, pp. 1227–1238, 1981.

[29] A. Der Kiureghian and P.-L. Liu, "Structural reliability under incomplete probability information," *Journal of Engineering Mechanics*, vol. 112, no. 1, pp. 85–104, 1986.

[30] A. F. Smith and A. E. Gelfand, "Bayesian statistics without tears: a sampling–resampling perspective," *The American Statistician*, vol. 46, no. 2, pp. 84–88, 1992.

[31] K. M. Zuev, J. L. Beck, S.-K. Au, and L. S. Katafygiotis, "Bayesian post-processor and other enhancements of Subset Simulation for estimating failure

probabilities in high dimensions," *Computers & structures*, vol. 92, pp. 283–296, 2012.

[32] I. Papaioannou, W. Betz, K. Zwirglmaier, and D. Straub, "MCMC algorithms for subset simulation," *Probabilistic Engineering Mechanics*, vol. 41, pp. 89–103, 2015.

[33] C. Andrieu and J. Thoms, "A tutorial on adaptive MCMC," *Statistics and Computing*, vol. 18, no. 4, pp. 343–373, 2008.

[34] J. L. Beck and S.-K. Au, "Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation," *Journal of Engineering Mechanics*, vol. 128, no. 4, pp. 380–391, 2002.